5. EXPERIMENTAL

The experimental procedure followed for the development of ACAT receptor structure and development of a pharmacophore based 3D-QSAR model is described in this section.

5.1. Development of models of ACAT receptors

The MBOAT conserved domain of SOAT1 has been constructed by using I-TASSER server. I-TASSER (Iterative Threading ASSEmbly Refinement)[157-159] is a hierarchical method for protein structure and function prediction. Structurally similar proteins' folds are first identified from the PDB by multiple threading approach LOMETS (LOcal MEta-Threading-Server). LOMETS is a locally installed PDB library for meta-threading approach containing multiple threading programs in I-TASSER server. Continuous fragments excised from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations with the threading of unaligned regions (mainly loops) built by *ab initio* modeling. Fragment assembly simulation is performed again starting from the SPICKER (near native model selection algorithm) cluster centroids, where the spatial restrains collected from the LOMETS templates and the PDB structures by TM-align are used to guide the simulations. The purpose of the second iteration is to remove the steric clash as well as to refine the global topology of the cluster centroids. The decoys generated in the second simulations are then clustered and the lowest energy structures are selected. The final full-atomic models are obtained by REMO which builds the atomic details from the selected I-TASSER decoys through optimization of the hydrogen-bonding network.

Initially the MBOAT conserved domain of the SOAT1 was identified by NCBI blast (**Figure 4** in result and discussion). The amino acid sequence which represents the MBOAT conserved domain of the SOAT1 (amino acids 286 to 520) was submitted to I-TASSER server. The conserved MBOAT domain of SOAT contains the putative active site residue.

The templates of the highest significance in the threading alignments were measured by the Z-score. Z-score is the difference between the raw and average scores in the unit of standard deviation. The top 10 templates that are selected from the LOMETS threading programs with Z-score for the MBOAT conserved domain of the SOAT1 are as follows.

Table 3: The top 10 templates selected from the LOMETS threading program.

Rank	PDB Hit	Iden1	Iden2	COV	Norm. Z-Score
1	3mktA	0.11	0.16	0.97	1.02
2	4cadC	0.09	0.18	0.93	1.04
3	3rceA	0.1	0.22	0.99	1.35
4	4wd7A	0.12	0.17	0.95	1.39
5	3wxvA	0.08	0.19	0.94	1.01
6	3oyrB	0.12	0.16	0.83	1.35
7	1r2fA	0.07	0.16	1	1.33
8	3d9sA	0.11	0.2	0.8	1.29
9	1oyzA	0.08	0.16	0.97	1.32
10	3ag3A	0.13	0.23	0.91	1.29

Ident1: is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
Ident2: is the percentage sequence identity of the whole template chains with query sequence.
COV: represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
Norm. Z-score: is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 means a good alignment and vice versa.

For each target, I-TASSER simulations generate tens of thousands of conformations (called decoys). To select the final models, I-TASSER uses SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and report up to five models which correspond to the five largest structure clusters. In Monte Carlo theory, the largest clusters correspond to the states of the largest partition function (or lowest free energy) and therefore have the highest confidence. The confidence of each model is quantitatively measured by C-score. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5 to 2], where a higher value of C-score signifies a model with a high confidence and vice-versa. TM-score and RMSD are estimated on the basis of C-score and protein length following the correlation observed between these qualities. The obtained values are shown below. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score. Although the first model has a higher C-score and a better quality in most cases, it is not unusual that the lower-rank models may also have a better quality than the higher-ranked models. If the I-TASSER

simulations converge, it is possible that less than 5 clusters are generated. This is usually an indication that the models have a good quality because of the converged simulations.

- C-score = -2.49
- C-score = -3.07
- C-score = -3.36
- C-score = -2.93
- C-score = -4.20

  Estimated TM-score = 0.42 ± 0.14

  Estimated RMSD = 11.4 ± 4.5Å

After the structure-assembly simulation, I-TASSER uses TM-align program to match the first I-TASSER model with all the structures in the PDB library. This section reports the top 10 proteins from the PDB which have the closest structural similarity (i.e. the highest TM-score) to the predicted I-TASSER model. Due to structural similarity, these proteins often perform similar function in the system.

Table 4: The top 10 proteins obtained with highest TM-Score.

Rank	PDB Hit	TM-Score	RMSD[a]	IDEN[a]	COV Alignment
1	4cadC	0.843	2.07	0.096	0.932
2	2ww9A	0.513	5.28	0.055	0.787
3	2wwbA	0.464	5.38	0.033	0.728
4	2wswA	0.462	5.84	0.058	0.791
5	3hfxA	0.461	5.76	0.076	0.783
6	2ix3B	0.461	5.48	0.034	0.757
7	3vvnA	0.46	6.07	0.036	0.766
8	5a63C	0.458	4.52	0.053	0.638
9	3p03A	0.455	5.79	0.026	0.783
10	4lz6A	0.455	6.07	0.065	0.77

Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
IDEN[a] is the percentage sequence identity in the structurally aligned region.
COV represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

As **4cadC** represents the integral membrane protease Rce1 and our target is also a membrane bound protein, Model1 was selected for the 3D structure of MBOAT conserved domain of the SOAT1. By using this obtained structure the 3D-structure of SOAT2 was developed with the help of MODELLER.[161]

5.2. Docking studies and development of pharmacophore based 3D-QSAR model

Molecular docking studies were performed using Glide (Schrodinger 2009). Glide performs grid-based ligand docking and explores favorable interactions between one or more small molecules i.e. ligands and a naturally larger receptor molecule, a protein. Glide provides three levels of docking precision: HTVS (high-throughput virtual screening); SP (standard precision) and XP (extra precision). The ligand structures were built within Maestro using the Build module and a single low energy conformation search was performed for molecules under study using OPLS_2005 force field at physiological *p*H condition using LigPrep module of Schrödinger. Docking calculations were executed in extra precision (XP) mode with the active sites of receptor (enzyme) structures.

To develop a pharmacophore based 3D-QSAR model, a set of 32 ACAT inhibitors synthesized in our laboratory were selected and the activity in IC_{50} (μM) was used for the development of pharmacophore model. The molar IC_{50} values were converted into log values as $1/\log IC_{50}$ (pIC_{50}) and were used as biological activity parameter in this study.

PHASE was used to develop pharmacophore and 3D-QSAR models. PHASE employs the tree-based partitioning algorithm and detects the spatial arrangements of functional groups thoroughly that are common and necessary for the biological activity for a set of high-affinity ligands. Activity data and a given hypothesis in combination create a 3D-QSAR model which should be able to recognize significant aspects of the molecular structure that govern the activity. All the 3D structures were drawn in MAESTRO and were minimized with default force field. In the beginning, the conformations for all the molecules were searched through a combination of Monte-Carlo multiple minimum (MCMM)/Low mode (LMOD) with maximum number of 250 conformers per structure and 100 steps of minimization. A pharmaset was defined by setting a threshold of $1/\log IC_{50} \geq 5.105$ for actives and $1/\log IC_{50} < 4.496$ for inactives. PHASE presents a standard set of six pharmacophore features, namely hydrogen bond acceptor (A), hydrogen bond

donor (D), hydrophobic (H) group, negatively ionizable (N) group, positively ionizable (P) group and aromatic ring (R). A five-point hypothesis was developed form the data set. Generated common pharmacophore hypotheses were checked by scoring alignment of the actives against a reference ligand by using default settings for 'Score Actives' to identify the pharmacophore from each box that resulted into the best alignment of the active ligands. The hypotheses were ranked according to the survival values for active and inactive compounds.

Further, the developed pharmacophore model was studied with an atom based 3-D QSAR model. In atom-based 3D-QSAR, a molecule is treated as a set of overlapping van der Waals spheres. Here van der Waals models of the aligned training set molecules were positioned in a regular grid of cubes, with each cube allotted zero or more 'bits' to account for the diverse types of atoms in the training set that occupy the cube. This illustration gives rise to binary-valued occupation patterns that can be used as independent variables to create partial least-squares (PLS) 3D-QSAR models. The data set with 32 molecules was divided as 26 molecules in training set and 6 into the test set and 3D-QSAR models were developed. The developed models were evaluated by cross validated R^2,Q^2, SD, RMSE and Pearson-*R*.